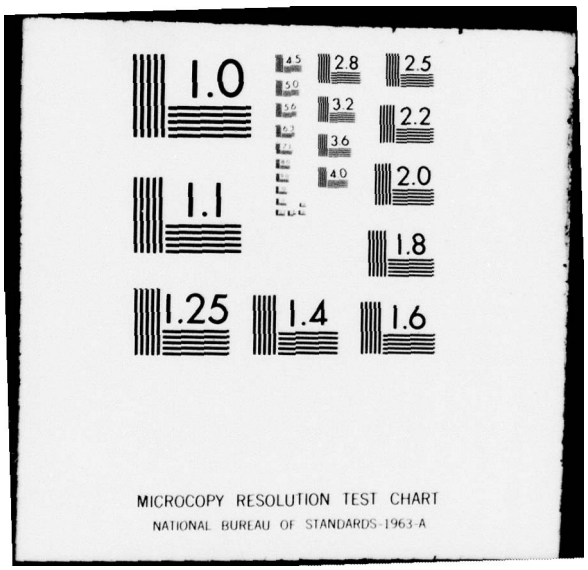
AD-A061 542 CONTROL ANALYSIS CORP PALO ALTO CALIF F/G 12/1
CONFIDENCE LIMITS FOR THE OPTIMUM OF A SIMULATION RESPONSE FUNC--ETC(U)
NOV 77 M A CRANE, D W ROBINSON N00014-72-C-0086
UNCLASSIFIED TR-86-24 NL

END
DATE FILMED
2-79
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
7.1
8.0
9.0
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

control analysis corporation

LEVEL II

12

PRINCIPALS AND OFFICERS

ROY P. HARVEY, President

Mr. Harvey has had broad experience in the application of computers to the solution of problems in industry and government, and particularly in the areas of management science and large-scale mathematical programming software systems. His prior associations are with Standard Oil Company of California, General Dynamics, and the British Scientific Civil Service.

MICHAEL A. CRANE, Senior Vice President

Dr. Crane has a broad background in health care delivery systems, stochastic modeling activities, large-scale simulation systems and transportation control systems. His prior associations include Westinghouse Corporation, International Business Machines, Inc., and Bell Telephone Laboratories.

GEORGE B. DANTZIG, Principal

Dr. Dantzig is Professor of Operations Research and Computer Science and the holder of the C. A. Criley Professorship in Transportation Science at Stanford University. He is also a member of the National Academy of Sciences and received the National Medal of Science from President Ford in 1976. Dr. Dantzig developed the simplex method for solving linear programs, and his textbook, **Linear Programming and Extensions** is considered the standard in the field. He has had extensive experience in the petroleum, chemical, steel, paper, banking, and airline industries in a consulting capacity.

DONALD L. IGLEHART, Principal

Dr. Iglehart, Professor of Operations Research, Stanford University, has made many contributions in the technology of inventory and queueing theory. He has consulted extensively in the fields of applied logistics, large scale simulation systems, error analysis, and program evaluation.

GERALD J. LIEBERMAN, Principal

Dr. Lieberman is Associate Dean of Humanities and Sciences and a Professor of Operations Research and Statistics, Stanford University. He is noted for his basic textbooks, **Introduction to Operations Research,** with F. Hillier, and **Engineering Statistics,** with A. H. Bowker. He has developed several standards and handbooks in the field of reliability and sampling inspection as well as engaging in consulting capacities in industrial and governmental operations.

RICHARD C. MOREY, Principal

Dr. Morey has had extensive experience in the areas of health care, energy utilization, capacity planning, inventory management, and budget allocation. His prior associations include Decision Studies Group Inc. as Vice President, Stanford Research Institute, Institute for Defense Analyses, and Eastman Kodak.

CONTROL ANALYSIS CORPORATION's independent auditors are Price Waterhouse and Company.

LEVEL II

12

CONTROL ANALYSIS CORPORATION

800 Welch Road

Palo Alto, California

9 Technical Report, No. 86-24

11 November 1977

14 TR-86-24

6 CONFIDENCE LIMITS FOR THE OPTIMUM OF A SIMULATION RESPONSE FUNCTION

12 32 p.

by

10 Michael A. Crane and David W. Robinson

15 This work was supported by the Office of Naval Research under contract N00014-72-C-0086 (NR-047-106).

407 383
78 11 21 036

TABLE OF CONTENTS

1. INTRODUCTION AND SUMMARY

This paper deals with stochastic simulation in which the system being studied can be controlled to some extent by setting the value of an input parameter. The system response for a given input will vary from simulation run to simulation run due to uncontrolled and random factors; we thus consider the average or expected response. This is usually called the simulation response surface. Since our work here involves only a single variable the term response curve might be more appropriate.

The simulation "response" will generally be identified with some measure of system effectiveness and the analyst will want to maximize (or possibly minimize) the response by choosing the optimum value for the input parameter. More formally, if we denote the input parameter by λ and the expected response by $g(\lambda)$, we wish to solve the problem

$$\text{maximize} \quad g(\lambda) \quad . \tag{1}$$

The main difficulty in solving (1) is that $g(\lambda)$ can be evaluated only approximately due to the random factors in the simulation. If a simulation is carried out with the input parameter λ fixed at, say, a_1, the output from the run will be a confidence interval for which

$$\Pr\{A_1^1 \leq g(a_1) \leq A_1^2\} = 1 - \gamma_1 \quad , \tag{2}$$

where γ_1 is a user-specified value and $[A_1^1 , A_1^2]$ is the confidence interval. Such intervals depend on some assumed probability distribution (often normal) for the simulation outputs. We do not consider this underlying distribution more explicitly in this paper; the reader may refer to

[2] or [3] for some examples. We do require that intervals of the form (2) be found for each simulation run that is carried out.

The basic result of this paper requires as given a set of confidence intervals $\{A_i\}$ for the response surface values at fixed input parameter values $\{a_i\}$. (There are a variety of approaches for obtaining these confidence intervals, e.g. batch means, independent replications, autoregressive schemes, regenerative approach, etc.) These are used to obtain confidence intervals for the solution λ^* to (1), the optimum value $g(\lambda^*)$ and a joint confidence region for the point $(\lambda^*, g(\lambda^*))$, in the case where the response surface is quadratic, i.e.

$$g(\lambda) = z_1 + z_2\lambda + z_3\lambda^2 \quad . \tag{3}$$

The technique applied here is related to a confidence band methodology developed in a number of previous papers ([2] - [4]). The basic idea is to find the locus of the optima of all possible quadratic functions which pass simultaneously through each of the confidence intervals. The result (under some mild convexity assumptions) is a compact region whose x-axis dimensions give the required confidence interval for λ^*. The probability coverage of the final region is found by multiplying together the probabilities of the individual response surface confidence intervals.

The main drawback to our method is that the probability coverage of the confidence interval for the optimum may be quite low when many different parameter settings are used. By restricting ourselves to a quadratic response surface with a single input parameter, then, we may obtain results with but three or four observations while keeping the probability coverage

high. Moreover, the experience of workers in nonlinear programming and approximation theory indicates that a quadratic approximation is in many cases an entirely adequate representation for a smooth function over a restricted range. Thus, the assumed form (3) for $g(\lambda)$ is not necessarily restrictive.

The major strength of our technique is that it does not depend on the distribution of the simulation outputs in any way. In fact, probabilities are used only to find the probability of the final region by a simple multiplication. There is no need for regression, for equal variance assumptions or for distribution theory in analyzing simulation outputs.

Most other approaches to finding the optimum of a simulation response function involve carrying out a search over the parameter space; see [5] for a review and evaluation of work in this area. In our method no explicit attempt is made to determine the parameter setting for the next simulation run; rather, the emphasis is on using the information already obtained to determine where the optimum might be. The best use of the present work, then, would be to carry out an analysis of the final results of a simulation, perhaps using only the last 3-5 parameter settings.

The remainder of the paper is organized as follows: Section 2 presents the basic results and Section 3 consists of a numerical example. The final section reports on some straightforward extensions to the basic results and applies them to the numerical example.

2. BASIC RESULTS

Suppose that the response surface $g(\lambda)$ is as given by (3) and that simulation runs have been performed for the input parameter settings $\lambda = a_i$, $1 \le i \le n$, where $a_1 < \ldots < a_n$, $n \ge 3$. Suppose further that the simulation results are

$$\begin{aligned} A_1^1 \le g(a_1) \le A_1^2 \quad &\text{with probability} \quad 1 - \gamma_1 \ , \\ A_2^1 \le g(a_2) \le A_2^2 \quad &\text{with probability} \quad 1 - \gamma_2 \ , \\ &\ldots \\ A_n^1 \le g(a_n) \le A_n^2 \quad &\text{with probability} \quad 1 - \gamma_n \ . \end{aligned} \tag{4}$$

Assuming the simulation runs are independent, $g(\lambda)$ will satisfy all of the inequalities (4) with probability $(1 - \gamma_1)(1 - \gamma_2) \cdots (1 - \gamma_n) = 1 - \gamma$. (Alternatively, one might deal with simulations with dependent results as long as the probability $1 - \gamma$ may be found; this seems unlikely in practice, however, so we will not consider it further here.)

With probability $1 - \gamma$, then, $g(\lambda)$ must pass through the points (a_i , y_i) , $i = 1, \ldots n$, where $A_i^1 \le y_i \le A_i^2$. Consider the vector $y = (y_1, y_2, \ldots , y_n)$; if three of the components of y are fixed, the corresponding value of $z = (z_1, z_2, z_3)$ from (3) may be found and the remaining components of y evaluated. Assuming without loss of generality that the three fixed components are y_1, y_2, y_3 we have

$$z = \frac{(y_1 , y_2 , y_3)}{(a_3 - a_1)(a_2 - a_1)(a_3 - a_2)} \begin{bmatrix} (a_3 - a_2)a_2a_3 & a_2^2 - a_3^2 & a_3 - a_2 \\ (a_1 - a_3)a_1a_3 & a_3^2 - a_1^2 & a_1 - a_3 \\ (a_2 - a_1)a_1a_2 & a_1^2 - a_2^2 & a_2 - a_1 \end{bmatrix} \quad (5)$$

Thus (5) gives the coefficients of a quadratic function passing through the points (a_i, y_i) , $i = 1, 2, 3$. Note that the components of z in (5) are simple linear functions of the components of y . We may thus refer to "$z_1(y)$" , say, to indicate the result of (5) for a specific choice of the y vector.

When there are more than three simulation runs it is likely that not all possible choices for the components of y will be admissible since there may be no quadratic function passing through the specified points. We thus require that y be a member of the constraint set Y , where

$$Y = \{y \mid A_i^1 \leq y_i \leq A_i^2 \;\; ; \;\; i = 1, 2, 3 \; ; \quad A_i^1 \leq z_1(y) + z_2(y)a_i + z_3(y)a_i^2 \leq A_i^2 \;\; ; \; 3 < i \leq n\} . \quad (6)$$

Note that all of the constraints in Y are linear.

The optimum of the quadratic function (3) occurs at the parameter setting

$$\hat{\lambda} = - z_2/2z_3 \quad . \quad (7)$$

This will be a maximum as long as $z_3 < 0$, i.e. the response surface is strictly concave. We assume for the present that, in fact, $z_3 < 0$; dealing

with linear or convex response surfaces is discussed more fully in Section 4.3.

The $\hat{\lambda}$ value given by (7) may be viewed as a function of y just as z_2 and z_3 are. Substituting from (5) and simplifying yields

$$\hat{\lambda}(y) = \frac{(a_3^2 - a_2^2)y_1 - (a_3^2 - a_1^2)y_2 + (a_2^2 - a_1^2)y_3}{2(a_3 - a_2)y_1 - 2(a_3 - a_1)y_2 + 2(a_2 - a_1)y_3} \quad . \tag{8}$$

Thus (8) gives the optimum $\hat{\lambda}$ of a quadratic function passing through the points (a_i, y_i), $i = 1, 2, 3$. Since we have assumed $z_3 < 0$, $\hat{\lambda}(y)$ is a continuous bounded function over the closed compact set Y. Thus, as a function of $y \in Y$, $\hat{\lambda}(y)$ takes on a maximum $(\bar{\lambda})$ and a minimum $(\underline{\lambda})$ and all intervening values.

Now, given that $g(\lambda)$ is in fact quadratic and that it satisfies (4) (i.e., is contained by all of the simulation output confidence intervals) the optimum input parameter must lie in the closed interval $[\bar{\lambda}, \underline{\lambda}]$, where $\bar{\lambda} = \sup_{y \in Y} \hat{\lambda}(y)$ and $\underline{\lambda} = \inf_{y \in Y} \hat{\lambda}(y)$. Since the probability that (4) holds is just $1 - \gamma$, we have

$$\Pr\{\underline{\lambda} \leq \lambda^* \leq \bar{\lambda}\} \leq 1 - \gamma \quad . \tag{9}$$

This is the desired confidence interval for λ^*. (The inequality results from the observation that the optimal parameter may lie in the given interval even in cases where the quadratic does not lie within the simulation output confidence interval.)

Finding $\underline{\lambda}$ and $\bar{\lambda}$ is quite straightforward. Examining the gradient vector found by differentiating (8):

$$\nabla_y \hat{\lambda}(y) = \frac{(a_2 - a_1)(a_3 - a_2)(a_3 - a_1)}{2[(a_3 - a_2)y_1 - (a_3 - a_1)y_2 + (a_2 - a_1)y_3]^2} \begin{bmatrix} y_3 - y_2 \\ y_1 - y_3 \\ y_2 - y_1 \end{bmatrix}$$

shows that the fraction will always be positive, since the denominator is proportional to z_3^2 . Thus, when two of the first three components of y are fixed the direction of change of $\hat{\lambda}(y)$ is constant. (The rate of change varies but not the direction.) $\hat{\lambda}(y)$ may thus be increased, say, by changing the unfixed y component until one of the constraints in (6) becomes binding; this shows that both $\underline{\lambda}$ and $\bar{\lambda}$ occur at extreme points of the set Y , i.e. three of the simulation output confidence interval inequalities (4) will be binding.

Now suppose that λ is fixed at some point $[\underline{\lambda}, \bar{\lambda}]$. We know that there is at least one y vector such that $\hat{\lambda}(y) = \lambda$, i.e. the quadratic function passing through (a_i , y_i) , $1 \leq i \leq n$, has its optimum at λ . In general there will be a set of such vectors and the maximum response value $g(\lambda)$ will vary as different y values are selected. Denote by $\hat{g}(\lambda ; y)$ this maximum and define the functions

$$\underline{g}(\lambda) = \inf_{\substack{\hat{\lambda}(y) = \lambda \\ y \in Y}} \hat{g}(\lambda ; y) \qquad (10)$$

$$\overline{g}(\lambda) = \sup_{\substack{\hat{\lambda}(y) = \lambda \\ y \in Y}} \hat{g}(\lambda ; y) \tag{11}$$

These functions then depict a joint confidence region for the point $(\lambda^*, g(\lambda^*))$, i.e.

$$\Pr\{(\lambda^*, g(\lambda^*)) \in G\} \geq 1 - \gamma \tag{12}$$

where

$$G = \{(\lambda, g) \mid \underline{\lambda} \leq \lambda \leq \overline{\lambda} \; ; \; \underline{g}(\lambda) \leq g \leq \overline{g}(\lambda)\} \; .$$

This follows just as for (9).

Now (10) and (11) are linear programs for fixed λ . We show this as follows: the first constraint $(\hat{\lambda}(y) = \lambda)$ may be rewritten as $-z_2(y) = 2\lambda z_3(y)$ according to (7); simplifying using (5) yields

$$y_2 = \frac{(a_3 - a_2)(a_3 + a_2 - 2\lambda)}{(a_3 - a_1)(a_3 + a_1 - 2\lambda)} y_1 + \frac{(a_2 - a_1)(a_2 + a_1 - 2\lambda)}{(a_3 - a_1)(a_3 + a_1 - 2\lambda)} y_3 \; . \tag{13}$$

(13) shows that as y_1 and y_3 change, y_2 shifts so as to keep the optimum fixed at λ . The value of $\hat{g}(\lambda ; y)$ is given by $z_1(y) - z_3(y)\lambda^2$; substituting from (5) and (13) gives

$$\hat{g}(\lambda\,;\,y) \;=\; \frac{(a_3-\lambda)^2}{(a_3-a_1)(a_3+a_1-2\lambda)}\,y_1 \;-\; \frac{(\lambda-a_1)^2}{(a_3-a_1)(a_3+a_1-2\lambda)}\,y_3 \qquad (14)$$

The solution to (10) and (11) will then have two components of y (say, y_1 and y_3) at their bounds, with the third component y_2 determined by the optimality constant $\hat{\lambda}(y) = \lambda$, or (13).

This analysis must be repeated when $a_3 + a_2 - 2\lambda = 0$, i.e. λ is at the midpoint of $[a_1, a_3]$. (Note that this point may not be in $[\underline{\lambda}, \bar{\lambda}]$.) In this case, we obtain

$$y_1 \;=\; y_3 \qquad (15)$$

$$\hat{g}\left(\frac{a_3+a_1}{2}\,;\,y\right) \;=\; y_1 + \frac{(a_3-a_1)^2}{4(a_3-a_2)(a_2-a_1)}\,(y_2-y_1)\;. \qquad (16)$$

There are two alternatives for computing the bound functions. If this is to be done by "brute force" (either manually or with a computer) it is suggested that the analyst select each possible set of two extreme points in y and investigate the behavior of $\hat{g}(\lambda\,;\,y)$ as a third component of y is varied and λ is adjusted to maintain optimality; this approach is feasible for perhaps three or four observations. For more observations, it is suggested that the simple linear programs (10) and (11) be solved directly for different values of λ.

Unfortunately, it is also tedious to obtain a confidence region for $g(\lambda^*)$ alone. The values

$$\underline{g} = \inf_{\lambda \in [\underline{\lambda}, \overline{\lambda}]} g(\lambda) \tag{17}$$

$$\overline{g} = \sup_{\lambda \in [\underline{\lambda}, \overline{\lambda}]} \overline{g}(\lambda) \tag{18}$$

will satisfy

$$\Pr\{\underline{g} \leq g(\lambda^*) \leq \overline{g}\} \geq 1 - \gamma \tag{19}$$

just as for (9) and (12). The simplest way to solve (17) and (18) is by inspection from the $\underline{g}(\lambda)$, $\overline{g}(\lambda)$ plots.

The values of $\underline{g}$ and $\overline{g}$ do not in general occur at an extreme point of y although they often seem to. It is suggested, therefore, that the $(\hat{\lambda}(y), \hat{g}(\hat{\lambda}(y); y))$ values be plotted for each feasible y vector with three or more of its components fixed at their upper or lower bounds. This will always yield the values for $\underline{\lambda}$ and $\overline{\lambda}$ and will often yield the values for $\underline{g}$ and $\overline{g}$.

3. NUMERICAL EXAMPLE

In this section we present a worked-out example as an aid to understanding the analysis of the preceding section. The following data are given:

i	a_i	A_i^1	A_i^2	γ_i
1	-1	1	2	.02
2	0	4	5	.02
3	1	1	4	.01

Thus, for example, $[1, 2]$ is a 98% confidence interval for $g(\lambda)$ at $\lambda = -1$. The probability content of the final region is $(.98)(.98)(.99) = .95$ approximately. The coefficients z_i are given by

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & -1 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} y_2 \\ \frac{1}{2}(y_3 - y_1) \\ \frac{1}{2}(y_1 - 2y_2 + y_3) \end{bmatrix} ,$$

while the function $\hat{\lambda}(y)$ is

$$\hat{\lambda}(y) = \frac{y_1 - y_3}{2y_1 - 4y_2 + 2y_3}$$

The set Y is given by $y = \{y \mid 1 \le y_1 \le 2 , 4 \le y_2 \le 5 , 1 \le y_3 \le 4\}$. The bounds on λ^* are then readily found by observing the extreme points of the set y :

$$\underline{\lambda} = -\frac{1}{10} \quad , \qquad \text{for} \qquad y = \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix}$$

$$\overline{\lambda} = \frac{1}{2} \quad , \qquad \text{for} \qquad y = \begin{bmatrix} \text{any} \\ 4 \\ 4 \end{bmatrix}$$

For any $\lambda \in \left[-\frac{1}{10} , \frac{1}{2}\right]$ and $y \in Y$, we have, for optimality,

$$y_2 = \frac{2\lambda - 1}{4\lambda} y_1 + \frac{2\lambda - 1}{4\lambda} y_3 \ ,$$

$$\hat{g}(\lambda ; y) = \frac{(1+\lambda)^2}{4\lambda} y_3 - \frac{(1-\lambda)^2}{4\lambda} y_1 \quad (\lambda \neq 0) \ ,$$

$$\hat{g}(0 ; y) = y_2 \ .$$

The resulting $\underline{g}(\lambda)$, $\overline{g}(\lambda)$ functions are obtained, in this case, by examining the values of $\hat{g}(\lambda ; y)$ when two components of y are set at the various extreme points and the third component is allowed to vary. Taking the minimum and maximum values for each λ yields:

$$g(\lambda) = \frac{2\lambda^2 + 8\lambda + 4}{2\lambda + 1} \quad -\frac{1}{10} \leq \lambda \leq \frac{1}{2} \quad (y_1 = 2 \ , \ y_2 = 4) \qquad (20)$$

$$\bar{g}(\lambda) = \begin{cases} \dfrac{\lambda^2 - 6\lambda + 1}{-4\lambda} & -\frac{1}{10} \le \lambda \le \frac{1}{14} \quad (y_1 = 2,\ y_3 = 1) \\ \dfrac{4\lambda^2 - 10\lambda + 5}{1 - 2\lambda} & -\frac{1}{14} \le \lambda \le 0 \quad (y_2 = 5,\ y_3 = 1) \\ \dfrac{4\lambda^2 + 10\lambda + 5}{1 + 2\lambda} & 0 \le \lambda \le \frac{3}{10} \quad (y_1 = 1,\ y_2 = 5) \\ \dfrac{3\lambda^2 + 10\lambda + 3}{4\lambda} & \frac{3}{10} \le \lambda \le \frac{1}{2} \quad (y_1 = 1,\ y_3 = 4) \end{cases} \tag{21}$$

The segments of each function are annotated with the values of the y vector components which remain fixed. The functions are plotted in Figure 1. Note the characteristic (and unusual) shape of the joint confidence region for $(\lambda^*, g(\lambda^*))$. Figure 1 shows the upper and lower confidence bands for the response $g(\lambda)$; see [3] for details.

By inspection we find that

$$\underline{g} = 4 \quad \text{for} \quad \lambda = 0$$

$$\bar{g} = 5.225 \quad \text{for} \quad \lambda = 3/10$$

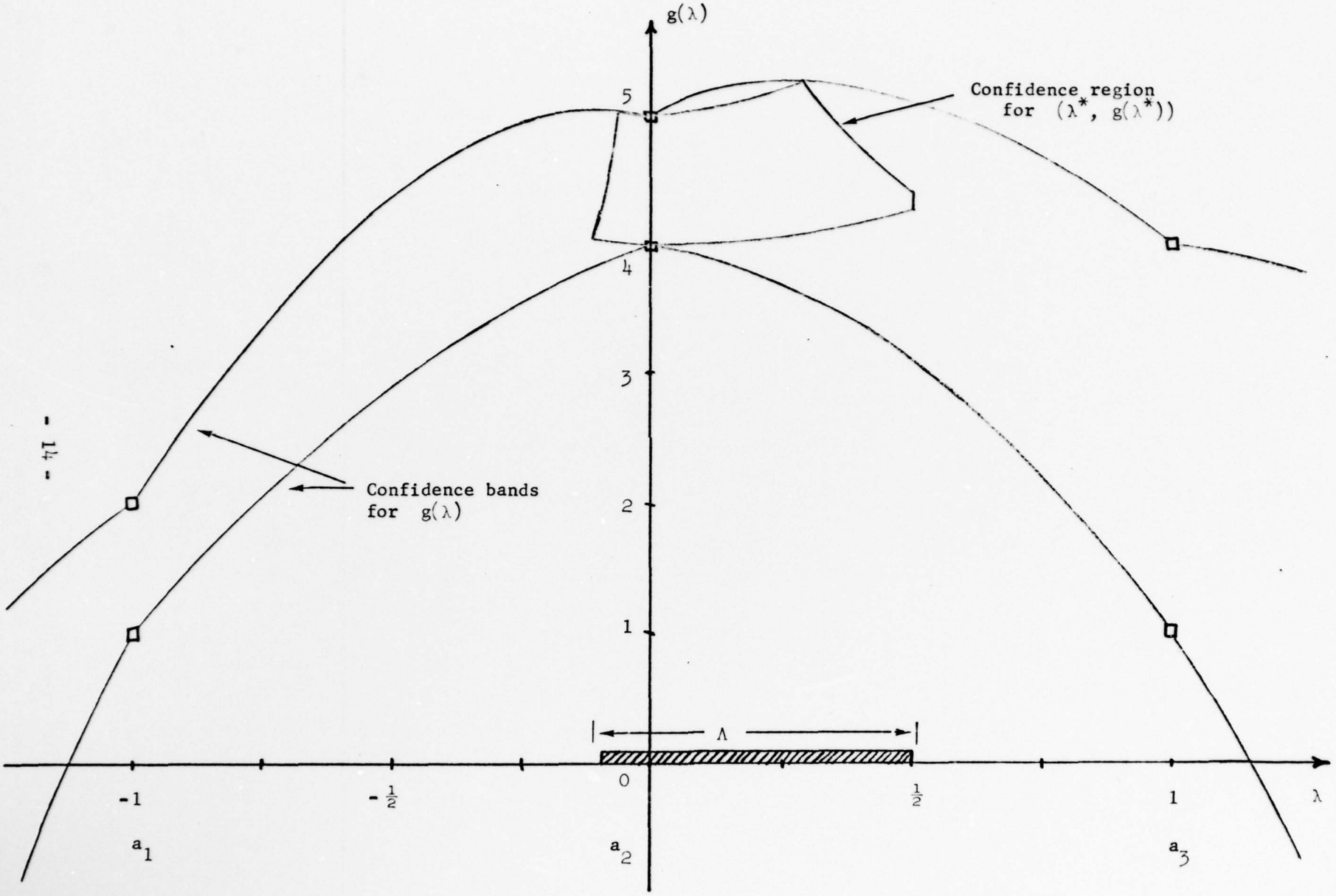


Figure 1. Confidence bands for $g(\lambda)$ and confidence region for $(\lambda^*, g(\lambda^*))$ for numerical example.

4. SOME EXTENSIONS

In this Section we take up a few obvious extensions to the work of Section 2. These are: dealing with a quadratic function with more than three simulation outputs, handling non-concavity and selection of good confidence interval widths at each base point. Extensions to the numerical example of Section 3 are also presented.

4.1 Additional Observations

Suppose that a fourth parameter setting $a_4 = 1/2$ is chosen for our numerical example and that, with probability .98, $4 \leq g(1/2) \leq 5$. Including this additional observation decreases the probability of our final results to about .93; this points up the major disadvantage of adding new observations to an existing solution.

The various optimization problems solved in Section 3 must be augmented by adding the constraints

$$4 \leq -\frac{1}{8} y_1 + \frac{3}{4} y_2 + \frac{3}{8} y_3 \leq 5 \quad .$$

When this is done, we obtain

$$\underline{\lambda} = -1/20 \quad \text{for} \quad y = \begin{bmatrix} 2 \\ 5 \\ 4/3 \\ 4 \end{bmatrix}$$

$$\overline{\lambda} = 1/2 \quad \text{for} \quad y = \begin{bmatrix} 2 \\ 4 \\ 4 \\ 4\ 1/4 \end{bmatrix}$$

The functions $\underline{g}(\lambda)$ and $\bar{g}(\lambda)$ can also be found; their form is similar to (20) and (21) and they are plotted in Figure 2. The results for the optimum value bounds are

$$\underline{g} = 4\ 1/2 \quad \text{for} \quad \lambda = 1/4 \ ,$$

$$\bar{g} = 5\ 1/6 \quad \text{for} \quad \lambda = 1/4 \ .$$

It is instructive to consider the reduction in size of the various confidence regions when the fourth observation is added. The intervals for λ^* and $g(\lambda^*)$ are reduced in length by 8.3% and 8.2%, respectively. By performing a somewhat tedious integration, one may find that the joint confidence region on $(\lambda^*, g(\lambda^*))$ is reduced in area by 24%.

4.2 Interval Width Selection

Instead of performing a simulation with an entirely new parameter setting as in the previous subsection, one may choose to perform more replications at one of the existing settings in order to narrow the confidence interval $[A_i^1, A_i^2]$ at that point. Note that this procedure leaves the probability content of the resulting confidence regions unchanged. Alternatively, it is usually possible to narrow any of the original confidence intervals if one is willing to accept the resulting decrease in the probability content of the final results. We thus consider the selection of a set of good widths for the $[A_i^1, A_i^2]$ intervals.

Because of the complex expressions that arise if one allows too much generality, it is difficult to give robust rules for choosing the interval

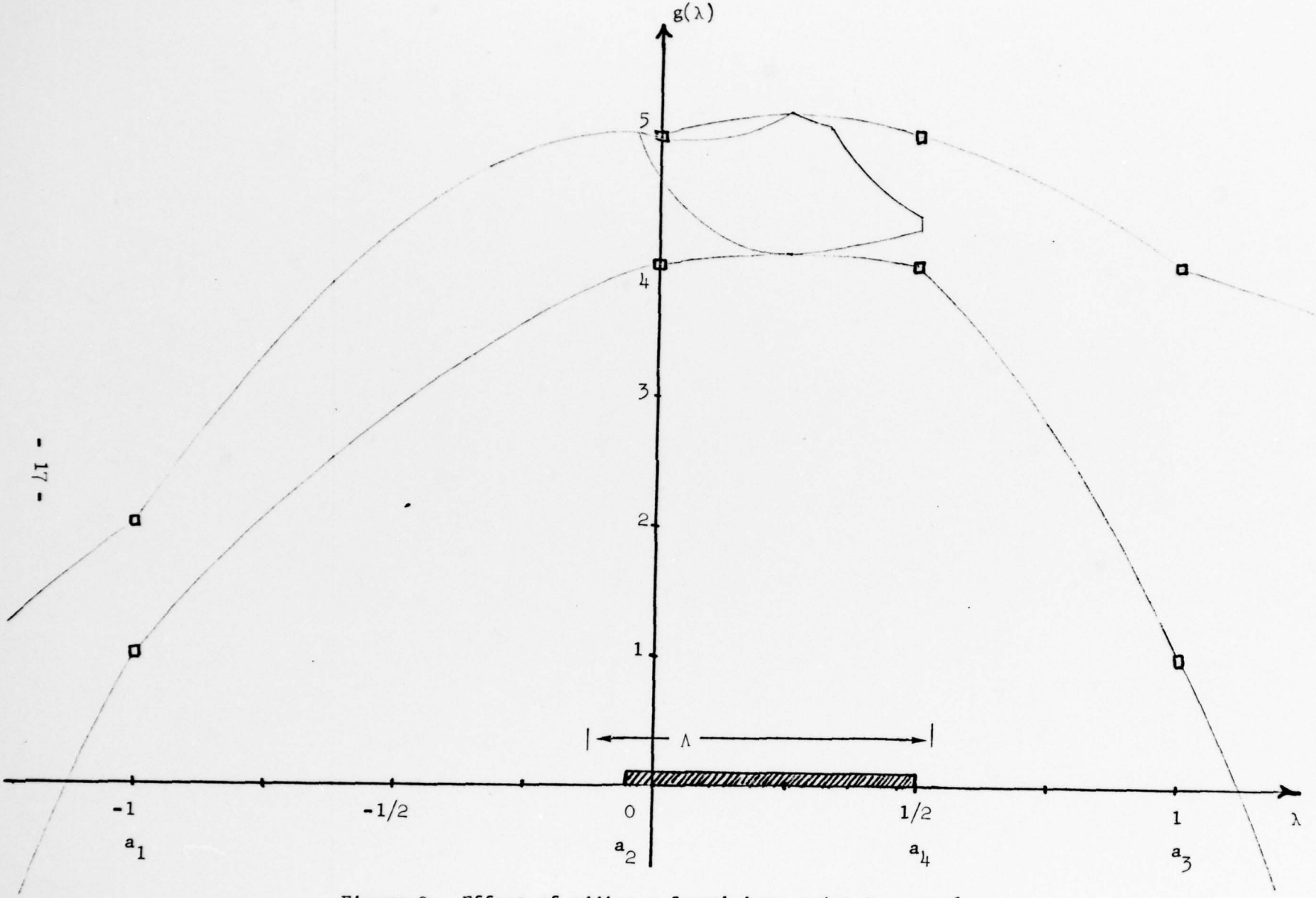


Figure 2. Effect of adding a fourth base point to example.

widths. Intuitively, it seems best to select all with about the same width, or else with about the same probability coverage. In order to consider this matter more concretely, we shall look more closely at the example of Section 3.

If the additional simulation replications carried out at $\lambda = a_4$ in Section 4.1 had been run instead for $\lambda = a_3$, the confidence interval $[A_3^1, A_3^2]$ would be based on twice as many observations and would thus be about 70.7% as wide as the original interval with the same probability coverage. The calculations of Section 3 were repeated, therefore, with $A_3^1 = 1.5$ and $A_3^2 = 3.5$; the results are plotted in Figure 3 and are summarized below:

$$\underline{\lambda} = \frac{1}{18} \qquad \text{for} \qquad y = \begin{bmatrix} 2 \\ 4 \\ 1\frac{1}{2} \end{bmatrix}$$

$$\overline{\lambda} = \frac{5}{14} \qquad \text{for} \qquad y = \begin{bmatrix} 1 \\ 4 \\ 3\frac{1}{2} \end{bmatrix}$$

$$\underline{g} = 4 \qquad \text{for} \qquad \lambda = 0$$

$$\overline{g} = 5\,{}^{25}/_{176} \qquad \text{for} \qquad \lambda = {}^{5}/_{22} \quad .$$

The size of $[\underline{\lambda}, \overline{\lambda}]$ thus decreases by 31% and of $[\underline{g}, \overline{g}]$ by 6.8%; the area of the $(\lambda^*, g(\lambda^*))$ region decreases by 41%. It may thus be concluded that in this instance it is preferable to narrow the third confidence interval than to take additional observations at some other point.

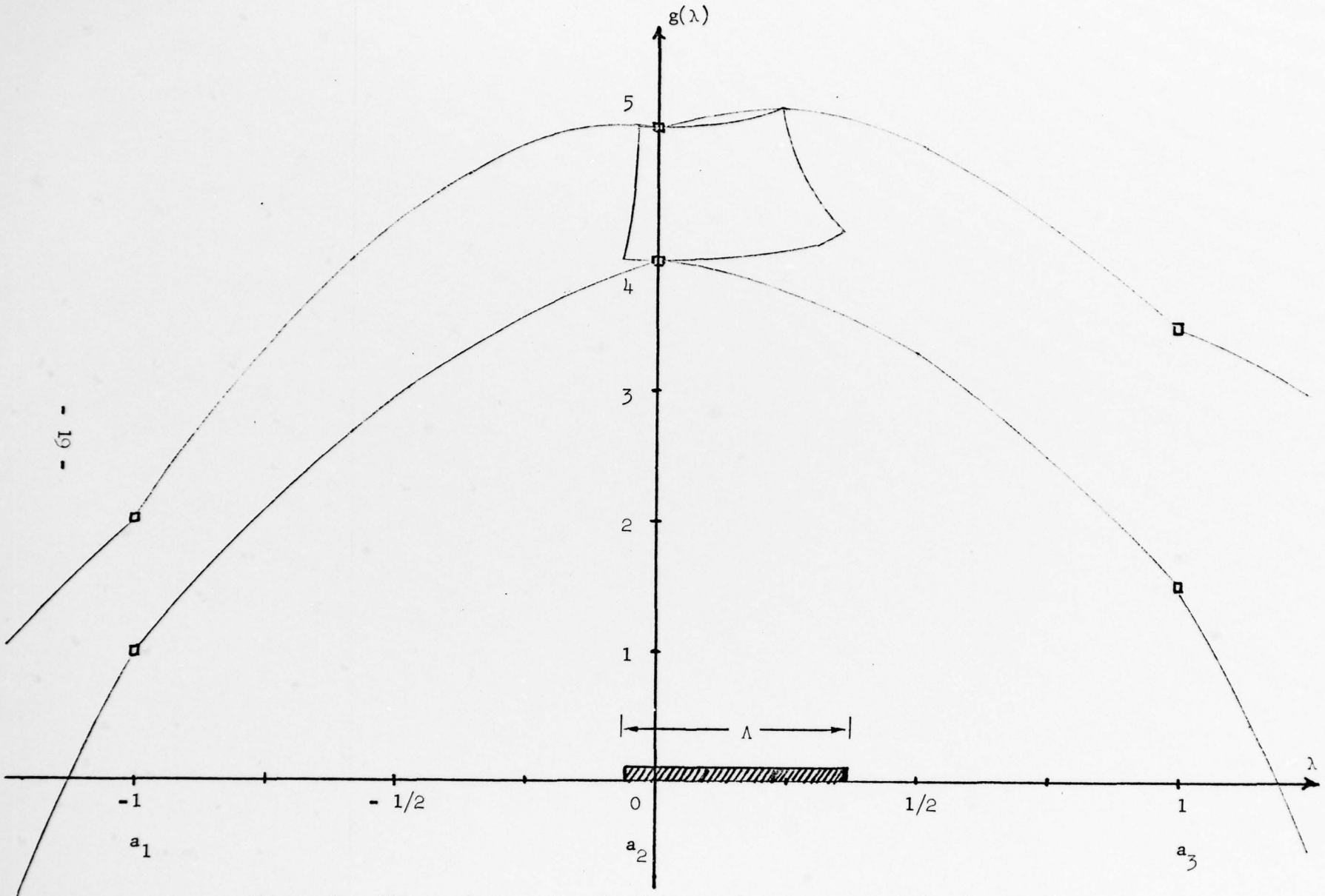


Figure 3. Effect of narrowing the third confidence interval in the example.

To pursue the matter further, we investigate how the widths of the $[\underline{\lambda}, \overline{\lambda}]$ and $[\underline{g}, \overline{g}]$ confidence intervals change as the $[A_3^1, A_3^2]$ interval changes. We assume that

$$A_3^1 = 2^{1/2} - \delta \quad , \qquad A_3^2 = 2^{1/2} + \delta \quad .$$

The resulting widths are plotted in Figure 4. Note that when $\delta = 3^{1/2}$ in this case, the vector $y = \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix}$ yields a linear function; the upper bound $\overline{\lambda}$ thus approaches $+\infty$ as δ increases.

If we further assume that the simulation response at $\lambda = a_3$ has a normal distribution (on an approximately normal distribution) we can plot the sizes of the two confidence intervals against their probability coverage; note that since the a_1 and a_2 intervals do not change in this case, the maximum probability is $(.98)^2 = .9604$. The result is shown in Figure 5. Note the sharp upward turn at a point corresponding to $1 - \gamma = .94$, i.e. $\gamma_3 \approx .02$.

Both Figures 4 and 5 lend some wieght to our previous observation that the "best" confidence regions for λ^* and $g(\lambda^*)$ result when the absolute widths of the simulation output confidence intervals are about the same. We must qualify this statement somewhat when the parameter settings are not evenly spaced, since a greater width is acceptable for parameter settings which are relatively distant from the optimum.

4.3 Non-Concave Functions

When the value of $z_3(y)$ found from (5) is not negative, the resulting $g(\lambda)$ function becomes unbounded as $\lambda \to \pm \infty$. As we saw in Section 4.2, this will cause either $\underline{\lambda}$ or $\overline{\lambda}$ or both to become unbounded as well.

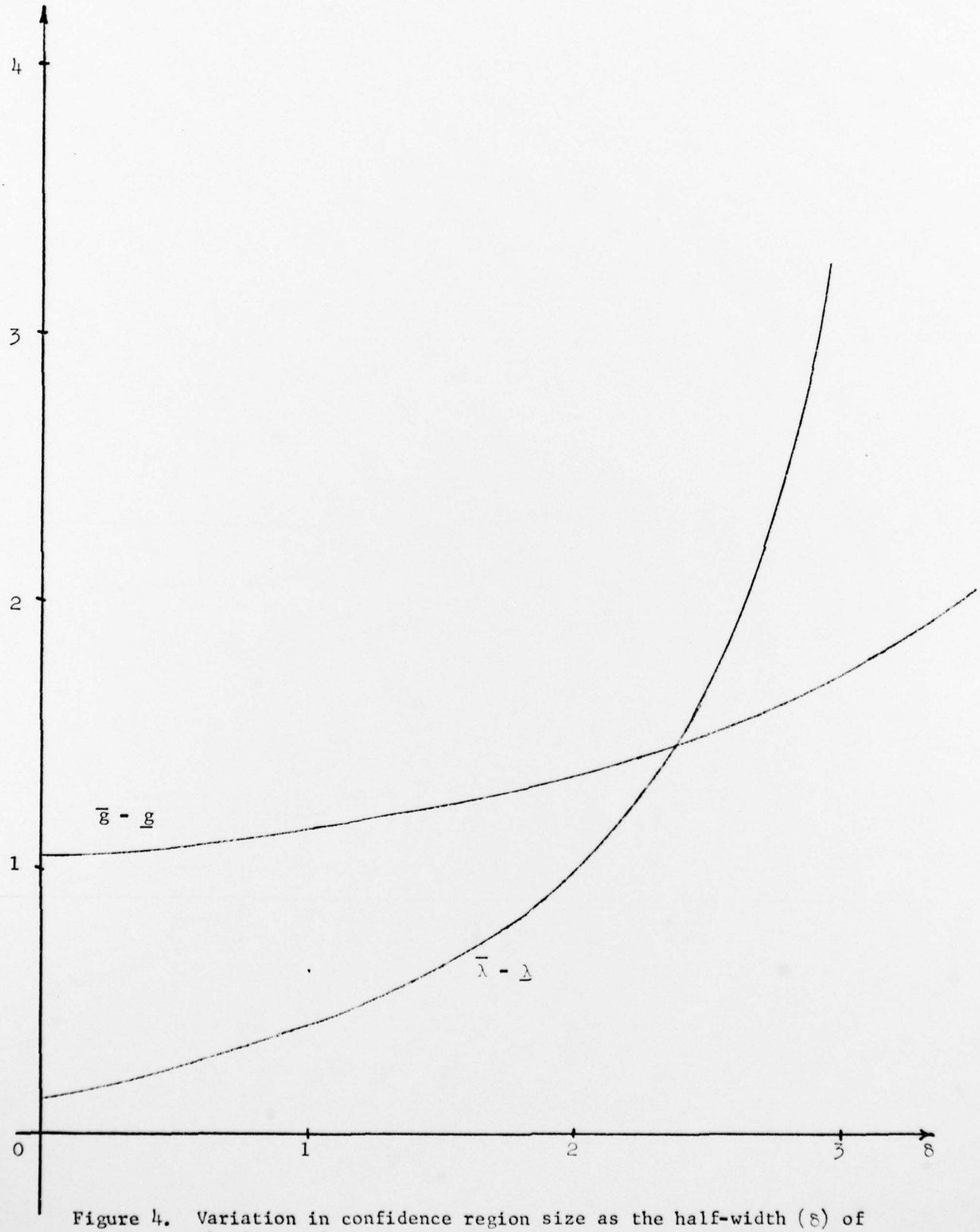


Figure 4. Variation in confidence region size as the half-width (δ) of the interval at $\lambda = a_3$ changes.

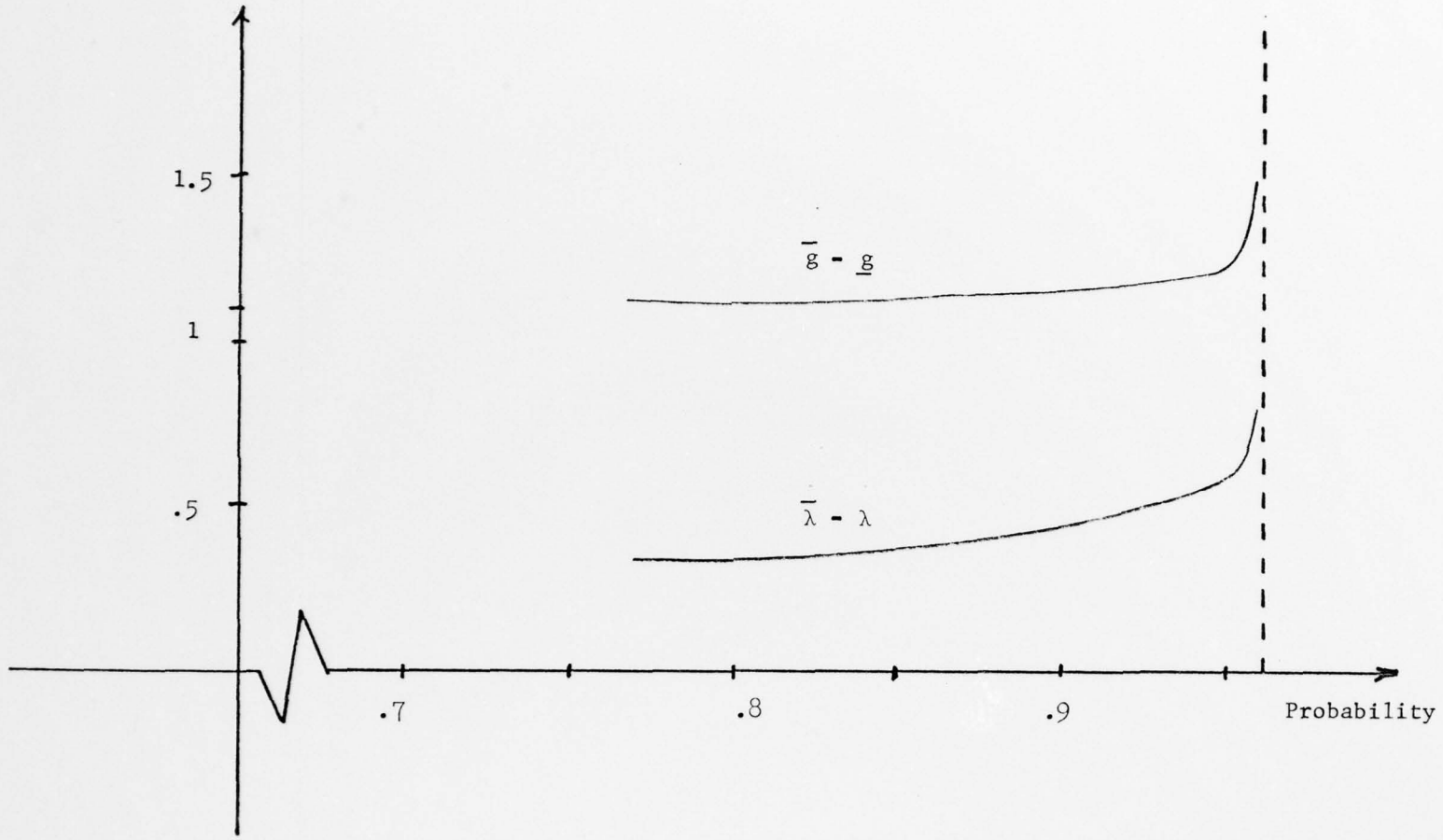


Figure 5. Region size vs. probability coverage when the observation at $\lambda = a_3$ is normally distributed.

There are two possible types of response to the situation when it turns out that $z_3(y) \geq 0$. The first response is based on the assumption that $g(\lambda)$ is actually a strictly concave function. In this case the analyst attempts to adjust the simulation results so as to obtain concave functions. The other type of response merely restricts the problem (1) to the closed interval $[a_0, a_{n+1}]$ where $a_0 \leq a_1$ and $a_{n+1} \geq a_n$ are user-specified constants. Of course, even a convex function attains a maximum over a closed bounded set. In the rest of this Section we discuss these two types of response in more detail.

The value of $z_3(y)$ will be negative as long as

$$A_j^1 > \frac{a_k - a_j}{a_k - a_i} A_i^2 + \frac{a_j - a_i}{a_k - a_i} A_k^2 \qquad (22)$$

for some $i < j < k$, $1 \leq i, j, k \leq n$. The adjustment process consists of modifying one triple A_i^2, A_j^1, A_k^2 if none of them satisfy (22). There are five ways to do this:

(1) By using the distribution theory that was used to obtain the original $[A_i^1, A_i^2]$ intervals, decrease the probability coverage (i.e., tighten the intervals) until (22) is satisfied. It may also be possible (if normal theory is applicable) to use asymmetric rather than symmetric intervals or to perform some other adjustement. It is likely that only a single interval will need to be altered.

(2) Carry out more simulation runs. These may be used to tighten the intervals already obtained or to investigate the response at an additional base point. Based on the results of this section, the

second alternative is suggested only when the responses at the existing points indicate that the optimum lies outside the current range.

(3) Inequality (22) may be expressed in terms of a constraint on y and explicitly introduced into the various optimization problems of Section 2 to insure that $z_3(y) < 0$. The main difficulty here is that (22) is a strict inequality; for this reason, a parameter $\mathscr{E}$ should be chosen and the following constraint used:

$$y_2 \geq \frac{a_3 - a_2}{a_3 - a_1} y_1 + \frac{a_2 - a_1}{a_3 - a_1} y_3 + \mathscr{E} \quad . \tag{23}$$

A good value for $\mathscr{E}$ would be a few percent of the maximum response obtained.

(4) The work of Barlow, et. al [1] on regression under order restrictions may be applied to adjust, say, A_1^2, A_2^1, and A_3^2 so that (22) always holds; this would involve estimates of the variability (i.e., variance) of the three values. This approach allows the analyst to use explicitly the additional information that $g(\lambda)$ is concave. This approach is somewhat involved and will not be discussed further here.

(5) The assumption that $g(\lambda)$ is concave may be questioned and a statistical test performed to determine its likelihood. This would involve information on the distribution of the simulation output response, however.

The constraint (23) offers a convenient means of carrying out the second approach, i.e. limiting the optimum λ^* to the interval $[a_0, a_{n+1}]$. In this case we solve for $\underline{\lambda}$ and $\overline{\lambda}$ as usual (the constraint protecting us from catastrophic unboundedness) and then set

$$\underline{\lambda} = \text{maximum } (a_0, \underline{\lambda})$$

$$\overline{\lambda} = \text{minimum } (a_{n+1}, \overline{\lambda}) \ .$$

As long as $\mathscr{E}$ is quite small this method should work well.

REFERENCES

[1] Barlow, P. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D., Statistical Inference under Order Restrictions, Wiley, 1972.

[2] Crane, M. A. and Iglehart, D. L., "Simulating Stable Stochastic Systems, I : General Multi-Server Queues," Journal of the ACM, V. 21, p. 103-113, 1974.

[3] Crane, M. A. and Iglehart, D. L., "Simulating Stable Stochastic Systems, II : Markov Chains," Journal of the ACM, V. 21, p. 114-123, 1974.

[4] Crane, M. A., "Techniques of Response Surface Estimation with Applications in Computer Simulation," Control Analysis Corporation Technical Report No. 86-19, August 1975.

[5] Farnell, W., McCall, C. H., Jr., and Russell, E. C., Optimization Techniques for Computerized Simulation Models, CACI Technical Report No. 1200-4-75, June 1975.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 86-24
2. GOVT ACCESSION NO.:
3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): Confidence Limits for the Optimum of a Simulation Response Function

5. TYPE OF REPORT & PERIOD COVERED: Technical Report

6. PERFORMING ORG. REPORT NUMBER: Technical Report No. 86-24

7. AUTHOR(s): Michael A. Crane and David W. Robinson

8. CONTRACT OR GRANT NUMBER(s): N00014-72-C-0086

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
CONTROL ANALYSIS CORPORATION
800 Welch Road Palo Alto, CA 94304

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR-047-106

11. CONTROLLING OFFICE NAME AND ADDRESS:

12. REPORT DATE: November 1977

13. NUMBER OF PAGES: 27

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office):
Distribution of this document is unlimited.

15. SECURITY CLASS. (of this report): Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Simulation, Response Surface Methodology, Statistical Analysis, Confidence Intervals, Optimization

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper sets forth a method for finding a confidence region for the input parameter setting λ^* for which a system being simulated attains its optimum expected response, $g(\lambda^*)$. A confidence region for $g(\lambda^*)$ and a joint region for $(\lambda^*, g(\lambda^*))$ are also developed. The present work is restricted to the case where $g(\lambda)$ is a quadratic function of a single variable.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

INSTRUCTIONS FOR PREPARATION OF REPORT DOCUMENTATION PAGE

RESPONSIBILITY. The controlling DoD office will be responsible for completion of the Report Documentation Page, DD Form 1473, in all technical reports prepared by or for DoD organizations.

CLASSIFICATION. Since this Report Documentation Page, DD Form 1473, is used in preparing announcements, bibliographies, and data banks, it should be unclassified if possible. If a classification is required, identify the classified items on the page by the appropriate symbol.

COMPLETION GUIDE

General. Make Blocks 1, 4, 5, 6, 7, 11, 13, 15, and 16 agree with the corresponding information on the report cover. Leave Blocks 2 and 3 blank.

Block 1. Report Number. Enter the unique alphanumeric report number shown on the cover.

Block 2. Government Accession No. Leave Blank. This space is for use by the Defense Documentation Center.

Block 3. Recipient's Catalog Number. Leave blank. This space is for the use of the report recipient to assist in future retrieval of the document.

Block 4. Title and Subtitle. Enter the title in all capital letters exactly as it appears on the publication. Titles should be unclassified whenever possible. Write out the English equivalent for Greek letters and mathematical symbols in the title (see *"Abstracting Scientific and Technical Reports of Defense-sponsored RDT/E,"AD-667 000*). If the report has a subtitle, this subtitle should follow the main title, be separated by a comma or semicolon if appropriate, and be initially capitalized. If a publication has a title in a foreign language, translate the title into English and follow the English translation with the title in the original language. Make every effort to simplify the title before publication.

Block 5. Type of Report and Period Covered. Indicate here whether report is interim, final, etc., and, if applicable, inclusive dates of period covered, such as the life of a contract covered in a final contractor report.

Block 6. Performing Organization Report Number. Only numbers other than the official report number shown in Block 1, such as series numbers for in-house reports or a contractor/grantee number assigned by him, will be placed in this space. If no such numbers are used, leave this space blank.

Block 7. Author(s). Include corresponding information from the report cover. Give the name(s) of the author(s) in conventional order *(for example, John R. Doe or, if author prefers, J. Robert Doe)*. In addition, list the affiliation of an author if it differs from that of the performing organization.

Block 8. Contract or Grant Number(s). For a contractor or grantee report, enter the complete contract or grant number(s) under which the work reported was accomplished. Leave blank in in-house reports.

Block 9. Performing Organization Name and Address. For in-house reports enter the name and address, including office symbol, of the performing activity. For contractor or grantee reports enter the name and address of the contractor or grantee who prepared the report and identify the appropriate corporate division, school, laboratory, etc., of the author. List city, state, and ZIP Code.

Block 10. Program Element, Project, Task Area, and Work Unit Numbers. Enter here the number code from the applicable Department of Defense form, such as the DD Form 1498, "Research and Technology Work Unit Summary" or the DD Form 1634, "Research and Development Planning Summary," which identifies the program element, project, task area, and work unit or equivalent under which the work was authorized.

Block 11. Controlling Office Name and Address. Enter the full, official name and address, including office symbol, of the controlling office. *(Equates to funding/sponsoring agency. For definition see DoD Directive 5200.20, "Distribution Statements on Technical Documents.")*

Block 12. Report Date. Enter here the day, month, and year or month and year as shown on the cover.

Block 13. Number of Pages. Enter the total number of pages.

Block 14. Monitoring Agency Name and Address *(if different from Controlling Office)*. For use when the controlling or funding office does not directly administer a project, contract, or grant, but delegates the administrative responsibility to another organization.

Blocks 15 & 15a. Security Classification of the Report: Declassification/Downgrading Schedule of the Report. Enter in 15 the highest classification of the report. If appropriate, enter in 15a the declassification/downgrading schedule of the report, using the abbreviations for declassification/downgrading schedules listed in paragraph 4-207 of DoD 5200.1-R.

Block 16. Distribution Statement of the Report. Insert here the applicable distribution statement of the report from DoD Directive 5200.20, "Distribution Statements on Technical Documents."

Block 17. Distribution Statement *(of the abstract entered in Block 20, if different from the distribution statement of the report)*. Insert here the applicable distribution statement of the abstract from DoD Directive 5200.20, "Distribution Statements on Technical Documents."

Block 18. Supplementary Notes. Enter information not included elsewhere but useful, such as: Prepared in cooperation with . . . Translation of *(or by)* . . . Presented at conference of . . . To be published in . . .

Block 19. Key Words. Select terms or short phrases that identify the principal subjects covered in the report, and are sufficiently specific and precise to be used as index entries for cataloging, conforming to standard terminology. The DoD "Thesaurus of Engineering and Scientific Terms" *(TEST)*, AD-672 000, can be helpful.

Block 20. Abstract. The abstract should be a brief *(not to exceed 200 words)* factual summary of the most significant information contained in the report. If possible, the abstract of a classified report should be unclassified and the abstract to an unclassified report should consist of publicly-releasable information. If the report contains a significant bibliography or literature survey, mention it here. For information on preparing abstracts see "Abstracting Scientific and Technical Reports of Defense-Sponsored RDT&E," AD-667 000.